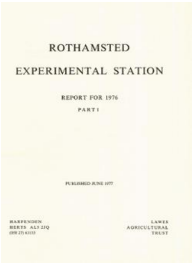
ROTHAMSTED RESEARCH

Report for 1976 - Part 1

Full Table of Content



Statistics Department

J. A. Nelder

J. A. Nelder (1977) *Statistics Department ;* Report For 1976 - Part 1, pp 325 - 340 **- DOI: https://doi.org/10.23637/ERADOC-1-133**

STATISTICS DEPARTMENT

J. A. NELDER

Staff

Head of Department J. A. Nelder, D.SC.

Senior Principal Scientific Officers
J. C. Gower, M.A.
R. H. Wimble, M.SC.

Principal Scientific Officers
B. M. Church, B.SC.
J. H. A. Dunwoody, B.A.
F. B. Leech, D.SC., M.R.C.V.S.
W. J. Lessells, B.SC.
G. J. S. Ross, B.A.
H. R. Simpson, M.A.

Senior Scientific Officers
N. G. Alvey, M.I.S.
C. F. Banfield, B.SC.
H. H. Spechter, PH.D.
A. J. Vernon, B.SC.

Higher Scientific Officers
C. J. Dyer, B.SC.
Mary G. Hills, B.A.
R. W. Payne, B.A.
J. G. Pearlman, M.SC.
K. Ryder, M.I.S., B.-TECH.
P. E. Sparrow, M.SC.
A. D. Todd, M.SC.
P. J. Zemroch, M.SC.

Scientific Officers
Mrs. Jill F. B. Altman, M.I.S.
R. J. Baker, M.SC.
Mrs. Linda C. Bassill, M.I.S.
P. W. Lane, B.A.
Penelope K. Leech
J. N. Perry, M.SC.
Janet Riley, M.SC.
R. P. White, B.SC.
J. Wood, M.SC.

Assistant Scientific Officers
Mrs. Janet I. Elsmere
Mrs. Pamela Tett

Secretary
Ruth T. Hunt

Clerical Staff
Mrs. Anne E. Arnold
Mrs. Jillian M. Bridgman
Mrs. Doris A. Court
Mrs. Elsie Davies
Mrs. Janet A. Day
Mrs. Merna M. Ellis
Mrs. Susan Greenland
Susan T. I. Hyde
Linda M. King
Moira Livings
Mrs. Elizabeth H. Mayton
Mrs. Patricia K. Neill
Mrs. Sheena Perry
Genevieve M. A. Rogers
Mrs. Brenda J. Watler
Janice R. Woods

Introduction

We are a Department with a large service component in our work, and the organisation of contacts with those whom we serve is of great importance. The establishment of links with other departments has been extended by a joint appointment with Physics Department for statistical work on their large micro-meteorological data base, and by the allocation of a post to collaborate with Botany Department on the development of mathematical models in plant physiology. The occupant of this post has previously specialised in control theory, which we hope may have useful insights to contribute to the development of models for self-organising systems.

Our work for the Agricultural Development and Advisory Service (ADAS) has now been formalised as a commission from the Ministry of Agriculture, Fisheries and Food, to provide services for the design, analysis, and interpretation of their experiments on both crops and livestock.

There has been a further increase in the use of our computer programs both inside and outside the Agricultural Research Service. Both Genstat and GLIM have been used as teaching aids, and it is becoming increasingly clear that computer programs can be a means of disseminating ideas comparable in power to the traditional scientific literature.

https://doi.org/10.23637/ERADOC-1-133

Practical applications

A complete catalogue of our consultative and collaborative work would be both long and tedious. The following paragraphs in this section are designed to exemplify collaboration with other departments at Rothamsted, with other Agricultural Research Council (ARC) Institutes, and with ADAS on both crop and livestock experiments. Other aspects include surveys and the considerable bulk of routine analysis, an essential service which underlies much of the work.

The population dynamics of the potato cyst-nematode. The current model (*Rothamsted Report for 1975*, Part 1, 322) was extended and a computer program written to simulate the changes in nematode number occurring under different crop-rotation policies when one or both British species were present. The nature of the stability of the equilibrium for a single species was examined and found to depend on the root damage caused by the nematode at its equilibrium level and the proportion of eggs laid which do not hatch in the first year. All three possible stability states predicted i.e. a steady approach to equilibrium, damped fluctuations about the equilibrium or wild undamped fluctuations, were found to occur in different species of cyst-nematode. The Lotka-Volterra competition model was fitted to data on final numbers of *Globodera rostochiensis* and *G. pallida* obtained after one generation for different initial proportions of the two species. The results suggested that under the conditions of the experiment *Globodera pallida* would always out-compete *Globodera rostochiensis* no matter how small their initial number. The experiment is being repeated with a better distribution of initial population sizes to confirm this result. (Kempton, with Jones, Nematology Department)

A multi-factor experiment on field beans. For the multi-disciplinary experiment on factors affecting yield of spring beans involving eight treatments each at two levels, a half replicate (128 plots) of the full 2^8 layout was used. Because irrigation was one of the treatments, and this could only be placed on whole plots, the design chosen was of eight blocks, each of two whole plots split into eight sub-plots.

Since all treatments were of equal importance the term chosen for aliasing was the eight-factor interaction. Main factors were therefore aliased with seven-factor interactions, etc. In such designs the residual error must be estimated from high-order interactions. In this experiment 28 degrees of freedom were available from the four-factor interactions. All main effects (apart from irrigation) and all two-factor interactions were estimable in the split plot stratum, but six of the three-factor interactions, involving the seven non-irrigation treatments symmetrically were not estimable. For some measurements sampling of all the plots was out of the question, and different experimenters had different requirements. This led to some interesting problems in deriving suitable sampling schemes. Various schemes were examined to allow half or one-quarter of the plots to be used. In these the three-factor interactions have to be used to estimate the error. Reasonable schemes were obtained by restricting observations to a particular level of one (or two) factors. If only irrigated plots were used, all main effect and two factor interactions would still be estimable in the split-plot stratum, but if a split-plot treatment were dropped, three two-factor interactions would not be estimable.

The results are encouraging in showing how the conduct of a combined experiment can be adjusted to meet the varying demands of experimenters in different disciplines. (Simpson, with Bardner, Briggs, Cockbain, J. M. Day, Fletcher, Legg, McEwen, Roughly, Salt, Webb and Witty)

Aphids on winter wheat. The data consist of seven weekly counts of three aphid species on 17 contrasting sites on a fifty-mile transect, together with concomitant observations

https://doi.org/10.23637/ERADOC-1-133

on degree of tillering, nitrate levels, degree of weediness, temperature, altitude etc. Two-way plots of log aphid counts against each variate have shown which variates may be clearly discarded. The next stage is to fit linear or other relationships where the graphs suggest that it will be profitable to do so. The most interesting correlations appear to be with nitrate levels and altitude, but the variable quality of the data may be obscuring other relationships. (Perry, with Henderson, Entomology Department)

Mildew sensitivity to ethirimol. This work continued with the analysis of the 1975 experiment on populations of *Erysiphe graminis*. It was found that LD90 values were the most sensitive for comparing the different populations. This is because even small doses of fungicide have a marked effect on preventing pustules forming—in some isolates, the smallest dose used decreased the number of pustules forming on test seedling leaves by 50% compared with the untreated controls. Using the LD90 estimates, it was possible to distinguish the different mildew populations, and to show that once mildew was well established on a plot, the sensitivity of that population did not change during the season. (Payne, with Bainbridge and Smith, Plant Pathology Department)

Automatic analysis of soil sections. Quantimet is an image-analysing computer, which is being used for determining the characteristics of soil sections. An initial investigation into sampling variation has been done. Here eight replicates (contained in tins) of a chosen horizon of two soils were sampled. Three thin sections were made from the soil in each tin, and photographs of each thin section were divided into nine frames by superimposing a 3×3 grid. Each frame was analysed by the Quantimet 720 to determine various characteristics of the soils: (i) porosity as measured by the proportion of the sampling area not containing material, (ii) the perimeter of the voids, which is a measure of the particle shapes and (iii) the number of voids. The distributions of these characters within soil sections and sampling tins were investigated. An hierarchical analysis of variance was obtained to estimate the variance components at the tin, section and frame levels of the experiment. These results will help to decide how much effort should be put into the various sampling levels (tins, sections, frames) when using the Quantimet in future investigations. (Banfield, with Murphy, Soil Survey)

Multivariate methods. Work continued on finding ways of classifying peat samples by measurements made in the laboratory so as to give good approximations to the classification made by soil surveyors in the field. A graph of rubbed fibre content plotted against absorbance of pyrophosphate extract was found to give reasonable results when subdivided into humic, mesic and fibric classes by diagonal lines. Principal components analysis was used to establish the best laboratory measurements for the separation of fibric from humic samples. (Banfield, with Bascomb, Soil Survey)

Canonical-variate analysis was used as a discrimination technique on four variables measured on 393 cyst nematodes obtained from 25 populations. The resulting representation of the population means plotted relative to the first two canonical variate axes (which accounted for 76% of the variation) showed very good separation between the Icelandic, European and South American populations. (Banfield, with Franco, Nematology)

The same procedure also worked well with data on 11 characters measured on 334 aphids from three populations. The populations, Gairi, Passeki and Phenax were found to separate well when plotted relative to the first two canonical variates. (Banfield, with Mr. H. Stroyan of the Plant Pathology Laboratory, Hatching Green)

Growth-invariant canonical-variate analysis was described in the *Rothamsted Report for 1975*, Part 1, 329. It was applied to carcase analysis where nine muscle measurements were made on each of a number of cattle sampled from three different breeds. The cattle

https://doi.org/10.23637/ERADOC-1-133

within each breed were known to be at different stages of growth from six months to two years of age. The growth-invariant technique adjusted the measurements so as to treat the cattle as all being at the same stage of growth and thus accentuated any differences between breeds. (Banfield, with Dr. H. J. H. MacFie, Meat Research Institute)

The technique was also used to separate several populations of each of two species of Paleocene planktic foraminifer. The fossilised organisms had been sampled at different depths of a borehole which were assumed to represent different stages of the species' evolution. Two methods of estimating the growth vectors within each sampled population, principal components and factor analysis, were tried, and for the two approaches the resulting canonical variates were very similar. The plotted growth-free canonical-variate means showed similar evolutionary changes for each of the two species considered. (Banfield, with Professor R. A. Reyment, University of Uppsala)

National Grassland Trial. Our assistance has been sought in summarising and interpretating the extensive data from the GM20 series carried out between 1970 and 1973 at 21 lowland sites throughout England and Wales. These tested six levels of nitrogen and three regimes of nitrogen application, and were given up to seven herbage cuts to simulate grazing.

The object of the trial was to study the extent to which soil, weather and climate influence the growth of grass and its response to fertiliser nitrogen, for the farmer the most effective management input for manipulating grassland yield within the limitations imposed by the environment. The data have now been collated and stored in a form suitable for easy use within Genstat. This will enable us to produce rapidly a technical report which will include all relevant information from each site.

Results indicate that differences of regime in terms of total production were negligible, with no detectable differences in form of the response. Previous optimal N rates had been calculated by assuming a quadratic response, but at 600 kgN ha^{-1} these were overestimates. Response to nitrogen, assuming a split-line form of response, is linear up to between 300 and 400 kgN ha^{-1} where maximum yield is attained. As current practice is to supply around 150 kg ha^{-1} to grassland, there is obvious room for improvement. Maximum yields vary from year to year and from site to site quite dramatically and while the latter variation may be attributed in part to N index more work needs to be done to relate the remaining variation to environmental factors.

The error distribution in the cut-by-cut analyses had originally been assumed to be Normal when untransformed. However, when checked it was found that a log transformation gave a better approximation to normality. Suitable curves have been found to describe the differences between cuts in their response to nitrogen, and work is in progress on the effect of the distribution of rainfall throughout the growing period, as well as its total amount. (Sparrow)

Livestock experiments. The design and analysis of ADAS experiments, including beef and dairy cattle, was again an important part of our work on livestock. This included the analysis of a complex dairy-cattle experiment involving 150 spring calvers with calving dates ranging over a three-month period. (Lessells and Altman)

Arrangements have been made to process all future data from the newly-created Feed Evaluation Unit at Drayton Experimental Husbandry Farm, the purpose of which is to develop predictive equations for estimating from laboratory determinations the extent to which forages can be metabolised. Standard forms for data recording have been designed. Two initial sets of data relating to silages and hay have been analysed. The silage results were the subject of a paper (with Mr. W. P. Barber and Mr. G. Alderman of ADAS)

https://doi.org/10.23637/ERADOC-1-133

presented to the Fourth Silage Conference at the Grassland Research Institute in September 1976. (Altman)

Considerable assistance was given to the Limousin and Simmental Breeds Steering Committee (MAFF) in producing their final report entitled 'The Evaluation of Limousin and Simmental Bulls in Great Britain', to be published by HMSO in January 1977. This involved the further analyses of the extensive data from the co-ordinated experiments done on five ADAS Experimental Husbandry Farms in 1973–74. (Lessells)

With Mr. S. C. Meadowcroft (ADAS) a paper is in preparation summarising the results of co-ordinated experiments testing the effects of nitrogen and stocking rate on beef cattle at grass. (Altman) Considerable progress has been made in producing a paper (with Mr. M. J. Strickland of ADAS and Dr. D. H. Broster of the National Institute for Research in Dairying) summarising the co-ordinated ADAS dairy cattle experiments done between 1965 and 1969. (Lessells)

We have kept closely in touch with the experimental programme at Gleadthorpe Experimental Husbandry Farm and have advised on plans for experiments both on egg production and on broiler chickens.

Experiments on egg production usually continue for one laying season and results are produced monthly for analysis and updating. This presents us with a very large amount of data that needs careful analysis because differences important to the industry can be quite small. Typically an experiment gives egg production data for 2000 hens for 12 months with concomitant data on egg quality and feed intake. Methods of analysing and presenting results from such experiments are being explored with the objectives of improving the efficiency of analysis and the intelligibility of summarised results. (F. B. Leech and Spechter)

An experiment to determine the effect of battery cage design on damage to eggs was analysed using the Rothamsted General Survey Program (RGSP). Fourteen different cage designs were tested during one laying season. The experiment showed clearly that damage varied with the construction of the cage floor in relation to the angle of tilt. A slack floor needed a steeper angle than a rigid floor, but too steep an angle increased damage. Pinhole cracks predominated in cages where eggs did not roll away satisfactorily; star and line cracks predominated when the eggs rolled away too quickly. (P. K. Leech)

We investigated the experimental evidence on the response of livestock to N applied to grazed grassland using data from 19 experiments done at four Experimental Husbandry Farms (EHF) over seven years. The results showed that, at current prices, N is profitable only for dairy cattle. This is in line with current practice as, at present, very little N is used on farms engaged in livestock rearing and fattening. However, the results also suggested that a comparatively small change in the price ratio between N and beef could lead to the use of N being profitable up to quite high levels, which could potentially result in much increased beef production. The results of this work have been written up. (Boyd and Wood)

Surveys. The largest survey we dealt with was again the Survey of Fertiliser Practice, followed by the National Milk Quality Survey.

Fertiliser practice. This survey is done predominantly to provide estimates of fertiliser consumption on individual crops; these are required by MAFF and the Fertiliser Manufacturers' Association (FMA) to establish trends and forecast future needs. In 1976, 600 visits were made by members of the ADAS Regional Soil Science departments, and 800 interviews by Farm Research Limited, on behalf of FMA.

Changes were made in the timing and organisation of the survey to obtain estimates by 1 October, two months earlier than in previous years, because the Price Review discus-

https://doi.org/10.23637/ERADOC-1-133

sions have been brought forward from February to the end of November. All the interviews made by Farm Research Limited, which formed a complete subsample covering England and Wales, were done in June and July, and a preliminary report based on these was available by the beginning of October. These preliminary estimates had to be based on farmers' forecasts of fertiliser applications up to the end of September, some of which were not fulfilled because of the drought. Most of the ADAS interviewing was done during August and September; this independent sample was analysed by the end of November and has been used to adjust the final estimate for the dry summer months. The figures show that use of P and K recovered in 1976 from the depressed levels of 1975 but the steady increase in use of N on grassland over recent years was checked by the exceptionally dry summer.

In an effort to increase information and accuracy without increasing costs (other than those due to inflation), the sampling method was modified in 1976 by stratifying by farm type as well as size, excluding 'agriculturally insignificant holdings' (<275 Standard Man-Days) and holdings of less than 50 acres crops and grass, and increasing the number of horticultural holdings in the sample.

Provision was made in the questionnaires for farmers to record areas and fertiliser use in either metric or conventional units; in the event only 35 out of 1380 farmers gave areas in hectares, and of these only three recorded fertiliser use in kilograms. The FMA request for estimates of brand shares also required substantial alterations to coding schedules and additional programming.

The mounting cost of such a survey increases the importance of collecting the most useful information and making it widely available, and a discussion meeting was held at Rothamsted in June at which staff of several ARC Institutes and of the MAFF Economics Department were present, as well as representatives of the fertiliser firms. The publication of our figures in Part 2 of the *Rothamsted Report* helps to make them readily available. (Church and Hills)

Milk quality. The analysis revealed some large differences between laboratories in their average results for some of the tests on milk samples. Although these effects can be eliminated when analysing this large set of data, they underline the need for regular monitoring of laboratories if comparable results are to be obtained in routine work. Indeed, results from a single unmonitored laboratory could, on occasions, be completely misleading. The tests were done at 25 laboratories and the milk quality results are being related to tests measuring the amount of bacterial contamination on milking equipment at the farms and to results of winter and summer surveys, at the same farms, of milking technique and of husbandry factors thought to affect the keeping quality of milk. Analysis of these complex inter-relationships is further complicated by the inter-laboratory differences noted above. (F. B. Leech and P. K. Leech)

Other surveys. Plans have been substantially completed for a national survey of grass weeds in 1977 for the National Wild Oats Advisory Programme. (Church, in collaboration with the Weed Research Organisation, ADAS and the Department of Agriculture for Scotland) Advice was given to the Potato Marketing Board about proposals for consumer preference studies, and to the Agricultural Training Board about training assessment (Church). The Central Veterinary Laboratory received advice on planning a National Mastitis Survey (F. B. Leech). Data on the survey of fertility in thoroughbred mares are now complete and validation checks have begun. (F. B. Leech and P. K. Leech)

Routine analysis. This year the total amount of data comprised 1·52 million items, about

https://doi.org/10.23637/ERADOC-1-133

the same as in 1974, and 12% less than in 1975. The average turn-round time was reduced by one day to nine days. Two continuing major commitments are the production of the yields of the field experiments for Rothamsted and Woburn in book form, and the analysis of ADAS field experiments, which this year numbered about 1000, a slight increase on last year.

The *Yields of the Field Experiments, 1975* (the report giving details and results of experiments supervised by the Field Plots Committee) was produced by a new and more efficient method.

With few exceptions, all tables of means and standard errors of differences, together with other calculated information—standard errors per plot, plot areas harvested, mean dry matter percentages etc.—were printed directly from suitably edited computer output, thereby saving typing time in copying and eliminating the need for further checking of the copy. Time was also saved by accumulating corrections at all stages on a single master copy of the details of each experiment (object, treatments, cultivations, notes etc.), leading to a single editing of a paper tape which was reproduced by an offset-litho process by printers as before.

For the 1976 edition, further improvements in presentation will be made, particularly regarding the uniformity of printed characters of the typed experimental details with the material produced by the computer. (Dunwoody, Dyer, Riley and Todd)

In addition to individual analysis of ADAS experiments, summaries of several series of such experiments were provided. These included:

(i) Correlation of soil analysis with crop response to fertilisers tabulated by region and N, P and K soil indices.
(ii) The warp soil series for potatoes in the East Midlands.
(iii) Nitrogen on winter wheat—a national series tabulated by soil series and N, P and K indices.
(iv) NPK manuring of potatoes ($\frac{1}{6}$ replicate of 6^3 experiments)—tabulated by regions and N, P and K indices.
(v) Comparison of phosphate fertilisers on grassland.
(vi) Fertiliser placement of early potatoes in Wales (summary over eight sites). (Dyer)

We again assisted with the analysis of tomato experiments from some of the ADAS Experimental Horticulture Stations. (Todd)

The service we provide for other Departments on the Station and for Broom's Barn was used by 108 workers during the year, involving data on a great variety of organisms. (Dunwoody, Riley and Todd)

Statistical programming

Genstat. Release 3·07 became available for System 4 and IBM 360/370 machines in January, and Release 3·08 followed in July for System 4, but the IBM version was delayed for several months because of difficulties of access to the 360 at Newcastle. Release 3·07 was converted for use on the Univac 1108 (Mr. R. Cormier, Statistical Research Service of Agriculture, Ottawa), the CDC 7600 (Dr. K. Y. Kwok, Manchester Regional Computer Centre) and the ICL 1900 series (Dr. P. Griffiths, Oxford University Computing Laboratory). Mr. B. P. Murphy adapted a CDC 7600 version for a CDC 6000 machine at the University of Western Australia; later Mr. Baghurst of the University of Adelaide became an official converter for the CDC 6000 series. The Nikola Poushkarov Institute of Soil Science in Bulgaria are to attempt a conversion for the Nova Eclipse range. We

https://doi.org/10.23637/ERADOC-1-133 *pp 8*

are most grateful for the work done by converters in making the program available on other machine ranges.

A further conversion was made in the Department for the ICL 4-75, a paged machine at the Edinburgh Regional Computer Centre (ERCC). This used the Forte compiler of Mr. G. E. Millard, and the whole of Genstat 3·07, except for the backing-store facilities, is now available. This is the first stage of a conversion for the new ICL 2900 range, on which the Forte compiler is also to be available. The help of Mr. Millard and other members of ERCC in this exercise is gratefully acknowledged. (Simpson)

With a large system like Genstat, improvements in efficiency need information on the relative amounts of use of different parts of the system by users. A special version of the program was loaded at Rothamsted for a week during November, which monitored the extent to which various groups of directives were used, and different diagnostics occurred. Genstat was itself used to produce summaries in the form of tables, and these will provide valuable evidence on the most intensively used parts of the system. (Simpson, with Beasley, Computer Department)

New facilities. Release 3.09 has been developed during the year and extensively tested at Rothamsted. It contains a major extension to the regression section, which now allows the fitting of generalised linear models. The original class of linear models with Normal errors is now widened to include, for example, log linear models for the analysis of counts, logit and probit models for percentages, and inverse polynomials with gamma errors for response surfaces. In addition the restricting of the models to quantitative variates and the main effects of qualitative variates has been lifted; interactions and mixed terms, where slopes vary with groups, can now be included. Genstat will thus contain the kernel of GLIM (Generalised Linear Interactive Modelling), the facilities of which have been widely appreciated. A form of summary has also been devised which provides, for a sequence of models with non-orthogonal data, an analogue of the standard analysis-of-variance table for balanced data. The presentation of the results of fitting models to non-orthogonal data is full of difficulties and experience with this new summary should help to show the most informative way of presenting such results to the user. (Lane, with Mr. R. Baxter, Division of Mathematics and Statistics, CSIRO, Australia)

Several new functions were provided for general calculations, including one to give the cumulative totals of a vector, and a general function to enable sets of elements of structures to be operated on. The Choleski decomposition of a positive-semi-definite symmetric matrix was also made available. (Alvey)

Improvements were made to the section on the analysis of designed experiments, including extra diagnostics to help users recover from reported errors. Some instances of partial aliasing can now be detected, and designs can be declared as known to be orthogonal, thus allowing punching errors in the design specification to be detected efficiently for all such orthogonal designs. (Payne)

For non-hierarchical classification two new criteria were added to the CLASSIFY directive. One criterion minimises the determinant of the within-class dispersion matrix and the other maximises the Mahalanobis distances between class means. (Banfield)

Numerous other revisions were also made to the system to improve efficiency and the detection of user errors etc. Again the number of faults in the system reported by users has been gratifyingly small.

Documentation. A Genstat newsletter, containing articles about the system, hints on its effective use, and special techniques, is now distributed twice a year along with updates to the standard documentation. (Alvey and Tett)

https://doi.org/10.23637/ERADOC-1-133

It was decided to combine the present set of user guides and the reference manual into a single document. An introductory chapter will outline the classes of problem for which the system can be used, and the existing user guides, extended and amended, follow. The present reference manual will become a technical appendix containing a formal description of all the facilities. A second draft of the new manual is now being circulated to selected people for comment, and it is proposed to issue it in 1977 with the first of a new series of releases of the program. (Alvey, Banfield, Gower, Lane, Nelder, Payne, Ross, Simpson and Todd)

Distribution. Currently 26 licences have been issued, 13 in the UK, four in the rest of Europe, four in North America and five in Australasia. Agreements on three further licences seem imminent, with others likely, following the recent availability of the ICL-1900 version. We have signed an agreement with Dr. E. W. Jones of the Computer Activity Group of Cornell University, USA, for the distribution of Genstat on IBM 360/370 machines in North America. (Alvey and Tett)

General Linear Interactive Modelling (GLIM). Release 2 of this program, which is distributed by the Numerical Algorithms Group (Oxford), is now available at more than 82 centres in 14 countries. It is extensively used within the Department, and by other ARS institutes. Substantial work has been done on Release 3, due to be available during 1977, including the provision of directives which allow branching and looping, the deletion of data, and better access to the various components resulting from the fit of a model. Information in the form either of GLIM instructions or of data can now be organised in subfiles easily accessible to the user, and calculations may be made on subsets of values of vectors, allowing, among other things, simple tabulations to be programmed. The user may define his own link functions and variance/mean relations, and may dump and recover the present state of the program, a facility useful when working in 'instant' batch mode.

The *User's Manual* has been extensively revised to include the new facilities, to expand the statistical section on the models and their fitting, and to remove obscurities. This work has been done in collaboration with Mr. M. R. B. Clarke (Queen Mary College, London), who is re-writing the algorithmic kernel for the fitting of generalised linear models. (Baker and Nelder)

Maximum Likelihood Program (MLP). The coordinates of maxima or minima of certain curves may now be printed, but otherwise no major facilities were added. Many internal improvements were made, and the code was modified to ease conversion problems. The manual was revised and enlarged, and course notes written. A regular timetable of releases has been defined and put into effect. (Ross)

An option was added to the directive FIT FREQUENCY to allow fitting of data on insect development in which the numbers of insects at each stage of development is observed at different sampling dates. The length of time an insect spends in each stage is assumed to be a gamma variate with index changing at each stage, and the death rate is assumed constant during the whole development period. A constant efficiency of sampling at each stage is assumed. (Kempton) Preliminary work has been done on generalising the MODEL directive and on including FIT SURFACE, and these extended facilities should be available in 1977.

MLP is now implemented at Bristol (ICL System 4), Newcastle, Cambridge and Paris (IBM 360/370), with a small ICL 1900 version at Barbados. Conversion is in progress at Sydney (CDC-Cyber 70 Series) and is about to start at Manchester (CDC 7600) and the London School of Hygiene (CDC 6600). (Ross)

https://doi.org/10.23637/ERADOC-1-133

Cluster Analysis Program (CLASP). A new facility was added to allow tests to have more than one value. Such tests must be coded as a group of binary tests, and an associated weight list indicates how the tests are to be grouped. The common blocks were reorganised so that jobs of any size may now be run subject to overall store availability. Many minor improvements were made, and the code was modified to ease conversions. The manual was improved and a prospectus prepared. CLASP is now implemented at Manchester and London (CDC 7600), Bristol (ICL System 4) with a small version at the British Museum of Natural History (IBM 1134). (Ross)

Genkey. Release 2 is now available on the CDC 7600 machines at Manchester and Canberra (CSIRO, Australia). Release 2.1 is now in preparation and contains extra facilities for keys printed in words; a page structure can be imposed and backward pointers can be printed which give, for each test, the index of the previous test in the key.

When a key is constructed using tests that are variable within species, identical groups of species may occur on different branches of the key, causing duplicate subkeys to be produced. A technique known as reticulation, which can be used to avoid printing such duplicate sections of the key, will be included in the new release. (Payne)

Data structures. Potential users of a statistical package should examine not only the algorithms it provides but should also investigate the types of data structures it supports and the extent to which output from algorithms can be stored internally in such structures; if such internal access to output from algorithms is available, a far wider range of analyses can be performed than those available from a single procedure in the package. An examination of ten popular packages revealed that some of the most widely-used are rather limited in this respect (Paper 21). (Payne and Nelder)

Other programs. A general algorithm has been written for transferring and swopping individuals between classes to optimise a given criterion. This provides a general framework within which specific algorithms for non-hierarchical classification may be embedded. (Banfield and Bassil) An algorithm to generate the exact distribution of the log-likelihood ratio statistic (deviance) in two-way tables has also been prepared. (Baker)

General regression work on survey data has been facilitated by a program which links RGSP and Genstat. Data on classifications are extracted from RGSP and recoded as dummy variates, and a published algorithm creates the sums-of-squares-and-products matrix which is passed to Genstat using a standard format. (F. B. Leech) Communication between packages via data in a standard form, usually without labels or captions, seems likely to be increasingly used, as it avoids the necessity of having to understand the internal programming conventions of each package.

The flexibility of the Genstat language has been found increasingly useful in constructing new algorithms. Thus the analysis of asymmetry as described under 'Theory' in this report was programmed in Genstat, as was a program for metric unfolding. (Banfield and Gower) While the analysis-of-variance algorithm in Genstat is essentially uni-variate, nevertheless the general facilities available have allowed multivariate analysis of variance (MANOVA) to be written as a macro, and this will be included in the macro library when testing is complete. (Banfield and Payne)

Theory

The relative abundance of species. The summarisation of extensive data such as that produced by the Rothamsted Insect Survey highlights the problem of describing species diversity in a quantitative way. A new diversity statistic $Q = S/2 \log(R_2/R_1)$ is proposed

https://doi.org/10.23637/ERADOC-1-133

for characterising species frequency distributions, where S is the total number of species and R_2, R_1 are the upper and lower quartiles of the distribution. Q depends on the numbers of species in the mid-range of abundance and can be shown to be better at discriminating between moth catches from different sites than either the information statistic or Simpson's index. It is not related to sample size.

Q showed smaller replicate variability when evaluated in terms of the estimated parameters of the log-series or log-normal models, even when the models did not fit the data particularly well and was fairly robust to choice of model. As a general rule the simplest model adequately describing the data should be used for estimating Q. The log-series model is satisfactory for species abundance of moth catches and the use of the parameter α, which is equivalent to Q for this model, is recommended. (Kempton)

Non-linear inference. The work reported last year (*Rothamsted Report for 1975*, Part 1, 329–330) has been further developed. Stable parameters may be stable *a priori* (i.e. largely independently of the particular data values) or *a posteriori* with the transformation from the original defining parameters depending on the data. Parameters are stable *a priori* if the log likelihood is sufficiently close to being a sum of squares for the purpose of rapid convergence of optimisation algorithms from reasonable starting values. Given the solution, further transformations may be sought to improve stability *a posteriori*. Stability may be improved in stages or continuously.

Confidence regions for stable parameters may transform back to irregular regions for defining parameters. If the transformation is complete and 1–1 then significance levels associated with the corresponding regions are the same for non-linear and for linear models, and this argument is supported by simulation. If the transformation is not 1–1, however, being partly ambiguous or undefined, the significance level depends on a measure of the region excluded by the transformation.

Limits for stable ordinate ratios were computed for several popular curves. Confidence regions were superimposed on the charts of feasible regions for each curve, and these show that for several four-parameter curves the coefficient of variation must be very small if the complete confidence region is to be feasible. These results illustrate the possible pitfalls in non-linear inference when the data convey too little information about the parameters in the model. (Ross)

Asymmetry in multivariate analysis. Many procedures in multivariate analysis depend on the notion of distance between points in a multi-dimensional space. A characteristic of distance is symmetry, i.e. the distance from A to B is the same as that from B to A. However, data frequently arises in conditions where assumptions of symmetry would be misplaced. Thus in tasting tests one compound may affect the subsequent ability of the taster to detect another, and do so asymmetrically. The switching of brands by consumers is also asymmetric, as are changes of state in longitudinal survey studies of animals.

The analysis of asymmetry, with the data in the form of an asymmetric matrix, proceeds by expressing the matrix as the sum of a skew-symmetric and a symmetric matrix. The symmetric part is subjected to principal-coordinate analysis, and the skew-symmetric part to a singular-value decomposition in the hope that a good approximation may be obtained using only a few of the singular values. This approach leads to the exploration of simply-structured skew-symmetric matrices, and some useful results have been obtained. The general procedure was applied to a model developed by Professor R. Curnow and Dr. Jorré of the University of Reading from which they had calculated the first-passage times for transitions from one amino acid of a protein to another. The transition times were asymmetric and analysis of these produced an interesting contour map, with a trough

https://doi.org/10.23637/ERADOC-1-133

at its centre. Transitions from high up the wall of the trough towards its centre are faster, and hence more frequent, than transitions in the reverse direction (Paper 9). (Gower)

Another application of asymmetry analysis concerns the relationship between pairs of objects classified by presence/absence characteristics. This can be summarised by four entries, a, b, c, d, say, in a two-by-two contingency table. The complement of the simple matching coefficient $b + c$ can be treated as a distance, and many objects can be represented approximately as points in few dimensions with inter-distances $(b + c)^{1/2}$. Can all the entries a, b, c, d be represented diagrammatically in a similar way? It turns out that $(b - c)$ can always be represented in one-dimensional form and its values used to label contours on the $(b + c)$ representation. The quantities $(a + b)$ and $(c + d)$ can be represented by including a zero and a unit point. Finally the quantities b and c may be represented by an asymmetry analysis. There is thus the possibility of representing (approximately) all the information in a set of 2×2 contingency tables, and the best ways of doing this are being investigated. (Gower and Banfield)

An interesting offshoot of this work arises from the relationship (first noticed by Cayley in the last century) between skew-symmetric and orthogonal matrices. Potential applications of the analysis of orthogonality are likely to be considerable, and as a first step, a study of simple orthogonal matrices has been completed. The results obtained are necessary for an understanding and interpretation of the canonical decompositions into elementary orthogonal matrices that arise in certain least-square approximations. They also throw light on and extend results of importance in numerical algorithms associated with eigenvalue and other linear algebra problems. (Gower, with Dr. A. G. Constantine, CSIRO, Australia)

Generalised linear models. For linear models with Normal errors, the residual sum of squares, which measures the goodness of fit, has a known distribution proportional to a χ^2 distribution. For other generalised linear models, the analogue of the residual sum of squares, a log-likelihood-ratio criterion called the deviance, does not have a known distribution, though it is known to be approximated by a χ^2 distribution as the amount of data becomes large. Little is known about the distribution when, for example, the data is of the presence/absence kind, i.e. each cell of a table contains either 0 or 1. Work has begun on approximations to the distribution of the deviance which can be used in judging the fit of generalised linear models. (Baker)

Commonwealth and Overseas

The continuing secondment of P. Walker throughout the year to CIMMYT (International Maize and Wheat Improvement Center) and the appointment of R. H. Wimble on 1 August 1976 to succeed Dr. Boyd has inevitably reduced the service which the Ministry of Overseas Development (ODM)-sponsored unit has been able to give during 1976. The work has continued to include overseas visits by staff, advice on the collection and interpretation of results from experiments and surveys, the statistical analysis of data on a variety of crops, and training for visiting workers.

For three weeks from mid-January, Wimble went to Malawi at the invitation of the Government. He visited the principal agricultural research station at Chitedze for a few days, but spent most of his time at the Biometrics Unit which is attached to Makoka Research Station near Zomba. Problems discussed related mainly to experimental design and analysis, and statistical software for the ICL 1901S machines available at Blantyre. A report on the visit was sent to ODM for transmission to the Government of Malawi.

In February Gilliver spent three weeks in Nepal advising on the design of experiments and the collection of data for the agronomy programmes at the British Gurkha Ex-

https://doi.org/10.23637/ERADOC-1-133

Servicemen Reintegration Scheme Agricultural Centres at Lumle and Pakhribas. Information was required urgently on suitable crops, with associated cultural practices, to be grown on terraces in mountainous terrain. The way in which the terraces are constructed leads to non-uniformity in the soil, with associated difficulties in the design of simple but accurate experiments. The strenuous itinerary included treks on foot of up to 30 miles to visit remote sites where cloud made the use of helicopters impossible. A report has been submitted with recommendations on general principles for experimental design in these conditions, and dealing also with the standardisation of recording procedures and future collaboration.

During the year work was completed on a new set of data-recording forms, intended in part for use in sending data to Rothamsted for analysis, but mainly for use in developing countries themselves after minor modification to suit local conditions. A number of clients were sent specimen sets and some useful comments were received. Copies were also given to a number of people visiting Rothamsted from overseas. (Gilliver and Wimble)

The following countries sent data for analysis during the year: Botswana, El Salvador, Fiji, Gambia, Kenya, Malawi, Malaysia, Nepal, Swaziland, Tanzania, Thailand, and Zambia. The crops concerned included beans, cassava, cocoa, cotton, grasses, groundnuts, legumes, maize, millet, oil palm, potato, sisal, sorghum, soya bean, sugar cane, sunflower, tobacco, and tomato. Some animal data were received from Kenya. Much of the analysis requested was routine, but the more substantial pieces of work included the following three items.

A paper has been published (Paper 14) on the results of a long-term rotational and manurial trial in Uganda, involving cotton, finger millet and groundnuts grown during the period 1936–64. The experiment, which ran for five cycles, involved a five-year rotation with three different resting periods, five types of resting cover, and farmyard manure at three levels. Yield trends differed from crop to crop, and a critical level of total soil nitrogen was suggested in partial explanation. Responses to farmyard manure in most crops increased over the years, with increasingly marked negative curvature in a pattern that showed most strongly in cotton crops immediately following the application of manure. (Wimble, with Mr. A. R. McWalter, Tate and Lyle Technical Services)

The analysis of a large number of District Fertiliser Trials in Swaziland has been completed. The aims of these trials included the determination of rates of nitrogen and phosphate dressings to be applied to six crops (maize, sunflower, cotton, potatoes, groundnuts and beans) in different areas in Swaziland defined by their soil properties. The trials were carried out over four seasons, 1969 to 1973, using approximately 300 sites of which a large proportion were used in consecutive years. The design used was systematic with four levels of nitrogen and phosphate, three levels of potash and one replication. Many sets of results had to be abandoned because of missing data or poor recording; this reduced the number of usable sites growing maize to approximately 60 for each season, those growing cotton to 20 and those growing any of the other four crops to under ten. Some 33 soils and location characteristics were recorded for each site, and for a number of sites the monthly rainfall figures and number of days with rain during the rainy season were also available.

Regression analysis indicated that only two variables, previous crop and November rainfall, had any consistent significant effects on responses of maize to N and P and that there was a large seasonal variation in response. As the November rainfall could not be used to determine fertiliser rates to be used in practice, the data was finally tabulated by year and previous crop, and for each combination, a double exponential curve with interaction was fitted to the N $\times$ P yield table. From these curves, recommended dressings for the whole of Swaziland were estimated by maximising net profit. The data from the other five crops did not justify more than simple tabulation by season. (Ryder and Wimble)

Two extensive genetic experiments on oil-palm progeny were done in Malaysia, using

https://doi.org/10.23637/ERADOC-1-133

cubic lattice-designs. The experiments tested 512 and 216 lines in three replicates.There were six variates measured for each experiment on two occasions. Within each plot there were six trees for the larger experiment and eight trees for the smaller, each recorded individually. The extent of the data was such that the time taken to punch and verify one replicate was approximately three days. Extensive use was made of the backing-store facilities in Genstat; first the plot means for each replicate were calculated and stored, and then recovered to form the raw material for analyses of variance. (Gilliver and Ryder)

One overseas worker from Sarawak, Miss Fatimah Othman, was attached to the Section for training during the summer for six weeks.

Staff and visiting workers

D. A. Boyd retired after 36 years in the Department. He had been in charge of the work we do for ADAS, and had made notable contributions to the design and interpretation of extensive experiments on fertilisers. He is succeeded by R. H. Wimble, who moves from the ODM section.

R. A. Kempton left to take charge of the Statistics Section of the Plant Breeding Institute, Cambridge. B. Gilliver left the ODM section and Janet Riley transferred to fill the vacancy. J. G. Pearlman, J. N. Perry and P. J. Zemroch were appointed. This last appointment is to a new post shared with the Physics Department.

J. A. Nelder attended the symposium 'Computer Science & Statistics: Ninth Annual Symposium on the Interface' held in Boston where he gave a paper (Paper 17); afterwards he visited Cornell and Pennsylvania Universities where he gave seminars. He and J. C. Gower attended the European Meeting of Statisticians held in Grenoble and both gave papers (Papers 18 and 9). J. C. Gower visited the Nikola Poushkarov Institute, Sofia where he acted as FAO consultant on a computerised soil management project review board and also advised on multivariate analysis and computing. He also attended the 'Second European Congress on Research into Odours (ECRO2)' held in Reading where he gave a paper. G. J. S. Ross was invited to the Conference on 'Mathematical Methods in Biology' held in East Berlin and organised jointly by the East German Society for Physical and Mathematical Biology and the East German Section of the Biometric Society. He also attended the 'Fourth International Statistical Conference of the Mathematical Institute of the Polish Academy of Sciences' held in Wisla, Poland. At both these conferences he gave papers. R. W. Payne attended the '9th International Biometric Conference' held at Boston, Mass., where he presented a paper by himself and Nelder (Paper 21). He also visited Union College, Schenectady, the Bell Telephone Laboratories at Murray Hill, and the National Institute of Health, Bethesda for consultations and discussions.

Genstat courses were given during the year by N. G. Alvey and H. R. Simpson in Southampton and Lancaster, and by C. F. Banfield and P. W. Lane in Edinburgh. In addition N. C. Alvey gave a course at Cornell University, Ithaca to help publicise the program in the United States.

Three temporary workers spent varying periods in the Department, two of them from overseas.

https://doi.org/10.23637/ERADOC-1-133

Publications

General Paper

1 Church, B. M. & Hills, M. G. (1976) Preliminary estimates of fertiliser use in England and Wales 1976: Comparison with previous years. London: *Ministry of Agriculture, Fisheries and Food* (SS/SAF/23), 3 pp.

Paper in Rothamsted Report, Part 2

2 Church, B. M. (1977) Use of fertilisers in England and Wales, 1976. *Rothamsted Experimental Station. Report for 1976*, Part 2, 189–193.

Research Papers

3 (Archer, F. C.), Victor, A. & Boyd, D. A. (1976) Fertilizer requirements of potatoes in the Vale of Eden. *Experimental Husbandry* No. 31, 72–79.

4 Banfield, C. F. (1976) Ultrametric distances for a single linkage dendrogram. *Applied Statistics* **25**, 313–315.

5 Banfield, C. F. & Bascomb, C. L. (1976) Variability in three areas of the Denchworth Soil Map Unit.II. Relationship between soil properties and similarities between profiles using laboratory measurements and field observations. *Journal of Soil Science* **27**, 438–450.

6 Boyd, D. A. & (Needham, P.) (1976) Factors governing the effective use of nitrogen. *Span* **19**, 68–70.

7 Boyd, D. A., Yuen, Lowsing, T. K. & (Needham, P.) (1976) Nitrogen requirement of cereals. Part 1. Response curves. *Journal of Agricultural Science, Cambridge* **87**, 149–162.

8 (Farrar, K.) & Boyd, D. A. (1976) Experiments on the manuring of maincrop potatoes on soils of Ross Series. *Experimental Husbandry* No. 31, 64–71.

9 Gower, J. C. (1977) The analysis of asymmetry and orthogonality. *Proceedings of the European Congress of Statisticians, 1976.*

10 Gower, J. C. (1976) Growth free canonical variates and generalized inverses. *Bulletin of the Geological Institutions of the University of Uppsala* **7**, 1–10.

11 (Johnson, P. A.), Boyd, D. A. & Sparrow, P. E. (1977) Manurial experiments with spring barley on chalk soils in southern England. *Experimental Husbandry* No. 32, 8–18.

12 Kempton, R. A. & Taylor, L. R. (1976) Models and statistics for species diversity. *Nature, London* **262**, 818–820.

13 Leech, F. B. (1976) Some biological factors affecting survival with particular relation to farm animals. *British Veterinary Journal* **132**, 335–345.

14 (McWalter, A. R.) & Wimble, R. H. (1976) A long-term rotational and manurial trial in Uganda. *Experimental Agriculture* **12**, 305–317.

15 (Needham, P.) & Boyd, D. A. (1976) Nitrogen requirement of cereals. Part 2. Multi-level nitrogen tests with spring barley in south-western England. *Journal of Agricultural Science* **87**, 163–170.

16 Nelder, J. A. (1976) An introduction to Genstat. *The Mathematical Scientist*, Supplement to **1**, No. 1, 57–58.

https://doi.org/10.23637/ERADOC-1-133

17 NELDER, J. A. (1977) Table operations in Genstat and RGSP. In: *Proceedings of the 9th Interface Symposium.* Massachusetts: Prindle, Weber & Schmidt, pp. 121–122.

18 NELDER, J. A. (1977) Intelligent programs, the next stage in statistical computing. In: *Proceedings of the European Congress of Statisticians, 1976.*

19 PAYNE, R. W. (1977) Some methods of identification. In: *Proceedings of 'Informatics 3'.* London: ASLIB.

20 PAYNE, R. W. (1977) Reticulation and other methods of reducing the size of printed diagnostic keys. *Journal of General Microbiology* **98**, 595–597.

21 PAYNE, R. W. & NELDER, J. A. (1976) Data structures in statistical computing. *Proceedings of the Ninth International Biometric Conference* **2**, 191–207.

22 (REYMENT, R. A.) & BANFIELD, C. F. (1976) Growth-free canonical variates applied to fossil foraminifers. *Bulletin of the Geological Institutions of the University of Uppsala* **7**, 11–21.

23 ROSS, G. J. S. (1975) Simple non-linear modelling for the general user. *Bulletin of the International Statistical Institute* **46**, 585–593.

24 (WEBBER, J.), BOYD, D. A. & VICTOR, A. (1976) Fertilizers for maincrop potatoes on Magnesian Limestone soils: experiments in Yorkshire, 1966–71. *Experimental Husbandry* No. 31, 80–90.

https://doi.org/10.23637/ERADOC-1-133